SEQUENCE AND STRUCTURE OF THE CORONAVIRUS PEPLOMER PROTEIN

R.J. de Groot[1], J.A. Lenstra[2], W. Luytjes[1], H.G.M. Niesters[1], M.C. Horzinek[1], B.A.M. van der Zeijst[2] and W.J.M. Spaan[1]

[1]Institute of Virology, P.O. Box 80.150 and [2]Section Bacteriology, P.O. Box 80.171, Veterinary Faculty, State University Utrecht, The Netherlands

INTRODUCTION

Coronaviruses display a characteristic fringe of large (17-20 nm), clubshaped peplomers, each consisting of a di- or trimer of the peplomer protein (Cavanagh et al. 1983). The peplomer protein, E2, plays an important role during the infection proces. It mediates the binding of virions to the host-cell receptors and is involved in membrane fusion. In addition, the E2 protein appears to be a major inducer of protective immunity to coronaviral infection (reviewed by Sturman and Holmes, 1983).

In the case of mouse hepatitis virus (MHV) and avian infectious bronchitis virus (IBV), the peplomer protein is synthesized as an N-glycosylated precursor of about 180K (reviewed by Siddell et al., 1982). This precursor is proteolytically cleaved, yielding two products of 80 to 90K, which remain noncovalently associated. Cleavage of the MHV peplomer protein is thought to be essential for the cell fusion activity (Sturman et al., 1985).

The peplomer proteins of the feline infectious peritonitis virus (FIPV) and the closely related porcine transmissible gastroenteritis virus (TGEV) differ from those of MHV and IBV in two respects. Firstly, proteolytic cleavage does not occur, although FIPV is fully capable to induce cell-fusion. Secondly, the peplomer proteins of FIPV and TGEV are larger, about 210K (Siddell et al., 1982; Boyle et al., 1984; Jacobs et al., 1986).

MHV, IBV and the FIP/TGE viruses are representatives of three separate antigenic clusters within the coronaviridae family (Siddell et al., 1983). Comparison of the nucleocapsid proteins of TGEV, MHV and IBV revealed a homology of about 27% in all three cases (Kapke and Brian, 1986), indicating a high degree of divergence.

In this report, we present a comparison of the peplomer proteins of IBV strain M41 (Niesters et al., 1986), MHV strain A59 and FIPV strain 79-1146. Because of the low overall homology (see below), the sequences that have been conserved in all three proteins are likely to be essential for common structural and/or functional features.

COMPARISON OF PEPLOMER AMINO ACID SEQUENCES

Cloning and sequencing of the peplomer genes of FIPV 79-1146 (de Groot et al, in prep.) and MHV A59 (Luytjes et al., in prep.) will be presented in detail elsewhere. The primary structure of the peplomer proteins was deduced from the nucleotide sequences. Apoproteins of 1162, 1324 and 1452 amino acid residues were predicted for IBV (Niesters et al., 1986), MHV and FIPV, respectively. At the N-terminal end of the deduced peplomer sequences a hydrophobic, 17-20 residue segment is found (Fig. 2). Binns et al. (1985) reported that this segment is absent in the mature peplomer protein of IBV. These findings suggest that the peplomer proteins are synthesized with a transient, N-terminal signal peptide (Wickner and Lodish, 1985). About 60 to 70 residues upstream of the C-terminal end, a distinct, hydrophobic region is found (Fig. 2), which most probably serves as a transmembrane anchor (Binns et al. 1985; Niesters et al., 1986).

As apparent from a Diagon comparison (Staden, 1982) of the amino acid sequences of the FIPV and MHV peplomer proteins (Fig. 1), most conserved residues are found in the C-terminal 60% of the protein.

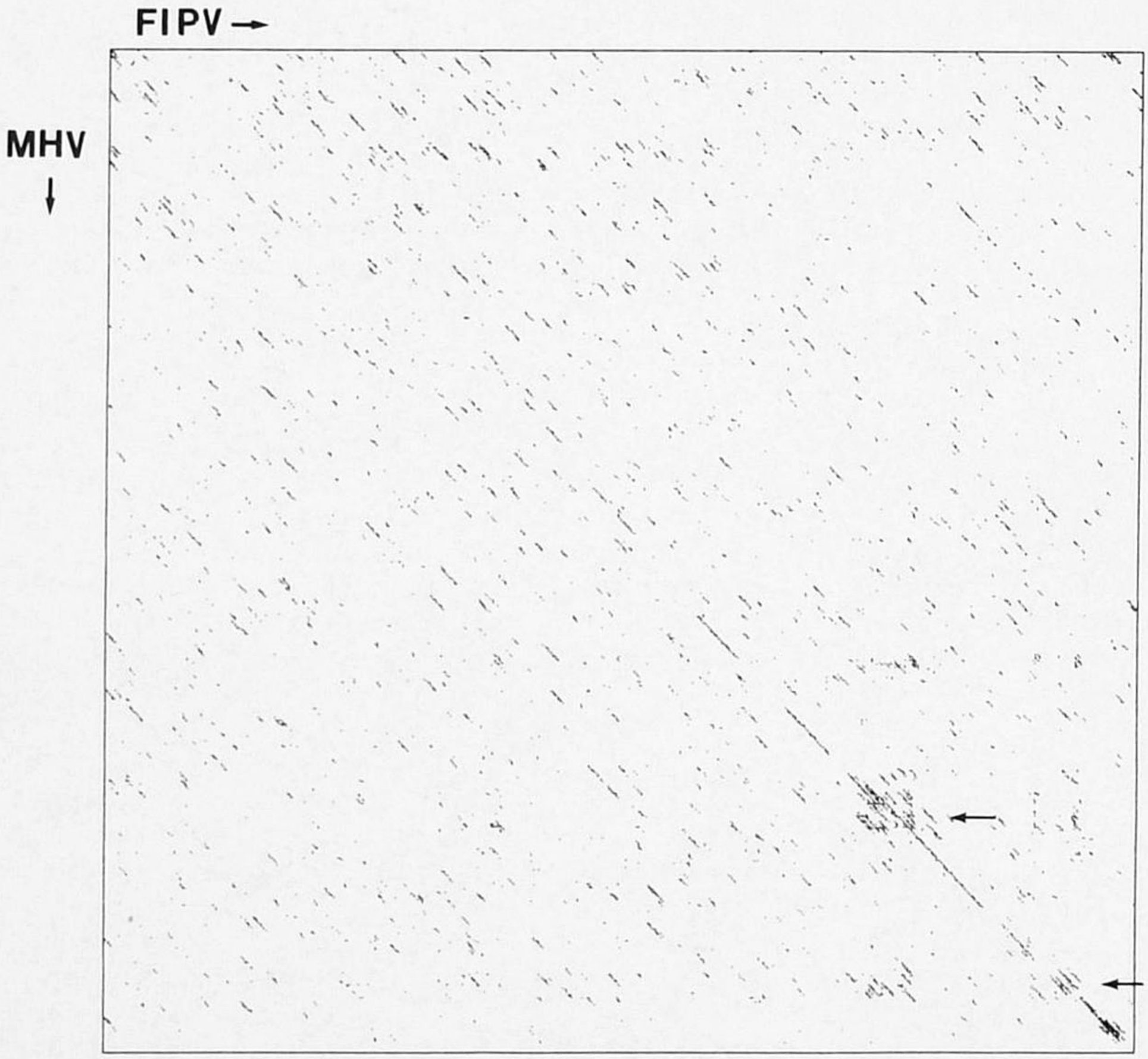

Fig. 1. Diagon comparison of the amino acid sequences of the FIPV and MHV peplomer proteins (proportional matching, span length: 21, minimal score: 221; Staden, 1982). The arrows indicate regions with an apparent repetitive character.

A more detailed comparison was made by combining FASTP alignments (Lipman and Pearson, 1985), Diagon plots and visual inspection (shown in Fig. 3). A schematic representation is given in Fig. 2.

Although there are some common sequence motives, the alignments in the N-terminal parts (residues 1-471 of FIPV, 1-398 and 652-724 of MHV, 1-206 of IBV) must be considered as tentative and may not be significant. In fact, apart from their supposed common origin, there is no real indication that the N-terminal domains, which presumably constitute the distal bulbous part of the peplomer, have the same three-dimensional structure. This is underlined by the distribution of the cysteine residues. In the C-terminal segments, corresponding to residue 689-1291 of FIPV, most cysteines are conserved in all three proteins. Their conservation indicates that they are probably involved in disulfide linkages important for the overall structure of the peplomer proteins. In contrast, there is no convincing matching of cysteines in the regions aligned to residues 1-471 of FIPV. Moreover, differences in the N-terminal regions account largely for the observed differences in molecular weights of the peplomer apoproteins.

As shown in Fig. 2, potential N-glycosylation sites are mainly found in regions of low homology. These sites appear to be particularly abundant in the N-terminal part of the protein and the low homology region immediately upstream of the transmembrane anchor. It may be noted that the peplomer protein of MHV contains less glycosylation sites than the peplomer proteins of FIPV and IBV (21 versus 35 and 29, respectively).

The presumptive transmembrane anchor is preceded by a highly conserved amino acid motive, KWPWYVWL, and followed by a peculiar, non-charged region, which is remarkably rich in cysteines (Fig. 3). We can only speculate about the function of these cysteine residues, e.g. acylation or membrane anchoring of the protein by disulfide bridges. However, the clustering of cysteines is clearly a typical feature of the coronavirus peplomer protein.

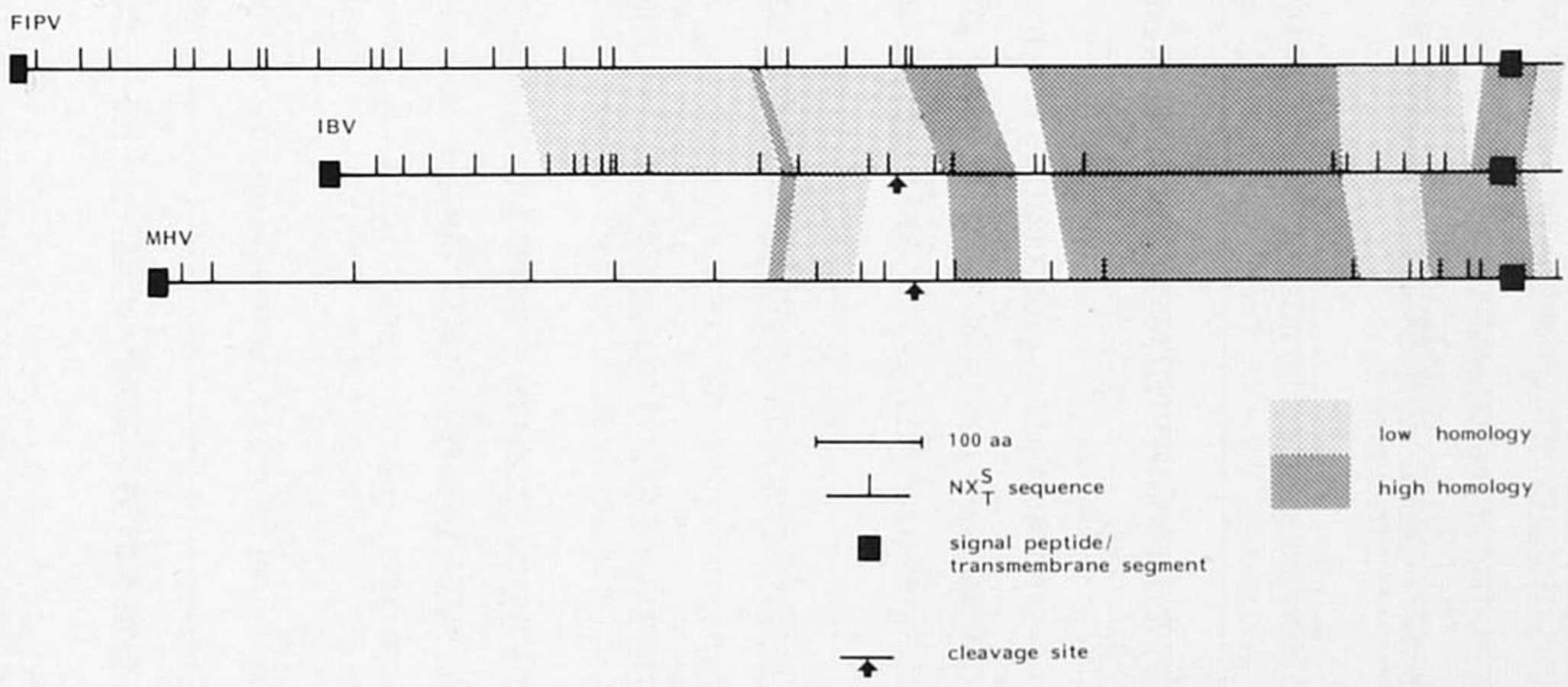

Fig. 2. Homology and potential glycosylation sites (NXS or NXT, except X=Pro) of peplomer sequences. Regions are considered highly homologous if two sequences are at least 30% identical.

```
FIPV: MIVLVTCLLLLCSYHTVLSTTNNECIQ-VNVTQLAGNENLIRDFLFSNFKEEGSVVVGGY   59
 IBV: M--LVTPLLLVTLLCVLCSAALYDSSSYVY----------------------------Y   29
 MHV: M--LFVFILFLPSC--LGYIGDFRCIQLVN-SNGANVSAPS----ISTETVEVSQGLGTY   51

FIPV: YPTEVWYNCSRTARTTAFQYFNNIHAFYFVMEAMENSTGNARGKP-LLFHVHGEPVSVII  118
 IBV: YQ-------------------------------------SAFRPPNGWHLHGGAYAVVN   51
 MHV: YVLDRVY-LNATLLLTGY-YPVDGSKFRNLALTGTNSVSLSWFQPPYLNQFNDGIFAKVQ  109

FIPV: ----SAYRDDVQQRPLLKHG-LVCITKNRHINYEQFTSNQWNSTCTGADRKIPFSVIPT-  172
 IBV: ISSESNNAGSSPGCIVGTIHGGRVVNASSIAMTAPSSGMAWSSS-QF--CTAH---CNFS  105
 MHV: NLKTSTPSGATAYFPTIVIGSLFGYT-SYTVVIEPYNGVIMASVCQYTICQLPYTDCKPN  168

FIPV: DNGTKIYGLEWNDDFVTAYISGRSYHLNINTNWFNNVTLLYSRSSTATWEYSAAYAYQGV  232
 IBV: DTTVFVTHCYKYDGCPITGMLQKNF-LRVSAMKNGQ---LFYNLTVSVAKYPTFKSFQCV  161
 MHV: TNGNKLIGFWHTDVKPPICVLKRNFTLNVNADAF-----YFHFYQHGGTFY-AYYADKPS  222

FIPV: S--NFTYYKLNNTNGLKTYELCEDYEHCTGYATNVFAPTSGGYIPDGFS--FNNWFLLTN  288
 IBV: N--NLTSVYLNGDLV--YTSNETTDVTSAGV----------------------------  188
 MHV: ATTFLFSVYI-GDILTQYYVLPFICNPTAGST---FAPRYWVTPLVKRQYLFNFNQKGVI  278

FIPV: SSTFVSGRFVTNQPLLINCLWPVPSFGVAAQEFCFEGAQFSQCNGVSLNNTVDVIRFNLN  348
 IBV: ---------------------------------------YFKAGGPITYK--VMRKVK-  205
 MHV: TSAVDCASSYTSEIKCKTQSM-LPSTGV-----------YELSGYTVQPVGVVYRRVAN  325

FIPV: FTADVQSGMGATVFSLNTTGGVILEISCYSDTVSESSSYSYGEIPFGITDGPRYCYVLYN  408
 MHV: LPACNIEEWLTARSVPSPLNWERKTFQNCNFNLSSLLRYVQAESLFCNNIDASKVYGRCF  385

FIPV: GTALKYLGTLPPSVKEIAISKWGHFYINGYNFFSTFPIGCISFNLTTGVSGAFWTIAYTS  468
 MHV: GSISVD----------------------------------------------------  391

FIPV: YTEALVQVENTAIKNVTYCNSHINN-IKCSQLTANLNNGFYPVASSEVGFVNKSVVLLPS  527
 IBV: ---ALAYFVNGTAQDVILCDGSPRGLLACQYNTGNFSDGFYPFINS--SLVKQKFIVYRE  260
 MHV: ----------------------------------------------------KFAVPRS  398

FIPV: FFTYTAVNITID----LGMKLSGYGQPIASTLSNI-TLPMQDNNTDVYCIRSNQFSVYVH  582
 IBV: NSVNTTFTLHN-----FTFHNETGANPNPSGVQNILTYQTQTAQSGYYNFNFSFLSSFVY  315
 MHV: RQVDLQLGNSGFLQTANYKIDTAATSCQLHYTLPKNNVTINNHNPSSWNRRYGFNDAGVF  458

FIPV: STCKSSLWDNIFNQDCTDVLEATAVIKTGTCPFSFDKLNNYLTFNKFCLSL---------  633
 IBV: KE--SNFMYGSYHPSCNFRLETINNGLWF----------NSLSVSIAYGPLQGGCKQSVF  363
 MHV: GKNQHDVVYAQQCFTVRSSYCPCAQPDIVSPCTTQTKPKSAFVNVGDHCEGLGVLEDN--  516

FIPV: ----------SPVGAN-CKFDVAARTRT--NEQVVRSLYVIYEEGDNIVGVPSDNSGL-H  679
 IBV: SGRATCCYAYSYGGPSLCKGVYSGELDL--NFECGLLVYVTKSGGSRIQTATEPPVITRH  421
 MHV: ------CGNADPHKGCICANNSFIGWSHDTCLVNDRCQIFANILLNGINSGTTCSTDLQL  570

FIPV: DLSVLHLDSCTDYNIYGRTGVGIIRR-T------NS--TLLSGLYYTSLSGDLLGFKNVS  730
 IBV: NYNNITLNTCVDYNIYGRTGQGFITNVTDSAVSYNYLADA--GLAILDTSGSIDIFVVQG  479
 MHV: PNTEVVTGICVKYDLYGITGQGVFKEVK--ADYYNSWQTLLYDVN-----GNLNGFRDLT  623

FIPV: D-GVIY-SVTPC-DVSAQAAVIDGAIVGAM--TSIN---SELLGLTHWTTTPNFYYYSIY  782
 IBV: EYGLTYYKVNPCEDVNQQFVVSGGKLVGIL--TSRNETGSQLL-----EN--QFYIKIT-  529
 MHV: TNK-TY-TIRSCYSGRVSAAFHKDAPEPALLYRNINCSYVFSNNISREENPLNYFDSYLG  681

FIPV: NYTSERTRGTAIDSNDVDCEPVITYSNIGVCKNGALVFINVTHSDGD-------------  827
 IBV: NGT-RRFRRS-ITENVANC-PYVSYGKFCIKPDGSIATIVPKQLEQF-------------  571
 MHV: CVVNADNRT---DEALPNCDLRMGAG-LCVDYSKSRRAHRSVSTGYRLTTFEPYTPMLVN  737

FIPV: --VQPI--STGNVTIPTNFTISVQVEYMQVYTTPVSIDCARYVCNGNPRCNKLLTQYVSA  885
 IBV: --VAPLLNVTENVLIPNSFNLTVTDEYIQTRMDKVQINCLQYVCGNSLDCRDLFQQYGPV  631
 MHV: DSVQSVDGLYE-MQIPTNFTIGHHEEFIQTRSPKVTIDCAAFVCGDNTACRQQLVEYGSF  796
```

```
FIPV: CQTIEQALAMGARLE-NMEVDSML-FVSENALKLASVEAF-NSTENLDPIYKEWPSIGGS   942
 IBV: CDNILSVVNSIGQKED-ME---LLNFYSSTKPAGFNTPFLSNVSTGEFNISLLLTT----   683
 MHV: CVNVNAILNEVNNLLDNMQLQVASALMQGVTISSRL-PDGISGPIDDINFSPLLGCIGST   855

FIPV: WLGGLKDILPSHNSKRKYGSAIEDLLFDKVVTSGLGTVDEDYKRCTGG--YDIADLVCAQ  1000
 IBV: ---------PSSPRRR---SFIEDLLFTSVESVGLPT-DDAYKNCTAGPLGFLKDLACAR   730
 MHV: CAEDGNG--PSAIRGR---SAIEDLLFDKVKLSDVGFV-EAYNNCTGGQEVR--DLLCVQ   907

FIPV: YYNGIMVLPGVANADKMTMYT----ASLA-GGITLGALGGGAVAIPFAVAVQARLNYVAL  1055
 IBV: EYNGLLVLPPIITAEMQTLYTSSLVASMAFGGIT------AAGAIPFATQLQARINHLGI   784
 MHV: SFNGIKVLPPVLSESQISGYTTG--ATAAAMFPPWS----AAAGVPFSLSVQYRINGLGV   961

FIPV: QTDVLNKNQQILANAFNQAIGNITQ-AFGKVNDAIHQTSQGLATVAKALAKVQDVVNTQG  1114
 IBV: TQSLLLKNQEKIAASFNKAIGR-MQEGF---------RSTSL-----ALQQIQDVVNKQS   829
 MHV: TMNVLSENQKMIASAFNNALGA-IQDGF-------------DAT-NSALGKIQSVVNAN-  1005

FIPV: Q-ALSHLTVQLQNNFQAISSSISDIYNRLDELSADAQVDRLITGRLTALNAFVSQTLTRQ  1173
 IBV: AI-LTETMASLNKNFGAISSVIQEIYQQLDAIQANAQVDRLITGRLSSLSVLASAK---Q   885
 MHV: AEALNNLLNQLSNRFGAISASLQEILTRLEAVEAKAQIDRLINGRLTALNAYISKQLSDS  1065

FIPV: AE-VRASRQ--LAKDKVNECVRSQSQRFGFCGNGTHLFSLANAAPNGMIFFHTVLLPTAY  1230
 IBV: AEHIRVSQQRELATQKINECVKSQSIRYSFCGNGRHVLTIPQNAPNGIVFIHFSYTPDSF   945
 MHV: TL-IKVSAAQ--AIEKVNECVKSQTTRINFCGNGNHILSLVQNAPYGLYFIHFSYVPISF  1122

FIPV: ETVTAWSGICASDGDRTFGLVVKDVQLTLFRNLDDKFYLTPRTMYQPRVATSSDFVQIEG  1290
 IBV: VNVTAIVGFCVKPANASQYAIVPANGRGIFIQVNGSYYITARDMYMPRAITAGDIVTLTS  1005
 MHV: TTANVSPGLCIS-GDR--GLAPKA---GYFVQDDGEWKFTGSSYYYPEPITDKNSVIMSS  1176

FIPV: CDVLF--VNATVIDLPSIIPDYIDINQTVQDILENYRP----------NWTVPEFTLD  1336
 IBV: CQANYVSVNKTVITTF-VDNDDFDFNDELSKWW-NDTKHELP---DFDKFNYTVP--ILD  1058
 MHV: CAVNYTKAPEVFLNTS--IPNPPDFKEELDKWFKNQTSIAPDLSLDFEKLNVT----LLD  1230

FIPV: IFNATYLNLTGEIDDLEFRSEKLHNTTVELAILIDNINNTLVNLEWLNRIETYVKWPWYV  1396
 IBV: IDSE-IDRIQGVIQGL----------NDSLIDLEKLSILKTYIKWPWYV  1096
 MHV: LTYE-MNRIQDAIKKL----------NESYINLKEVGTYEMYVKWPWYV  1267

FIPV: WLLIGLVVVFCIPLLLFCCFSTGCCGCIGCLGSCCHSICSRRQFENYEPIEKVHVH      1452
 IBV: WLAIAFATIIFILILGWVFFMTGCCGCCCGCFGIMPLMSKCGKKSSYYTTFDNDVVTEQN  1056
 MHV: WLLIGLAGVAVCVLLFFICCCTGCGSCCFKKCGNCCDEYGGHQDSIVIHNISSHED      1324

 IBV: RPKKSV                                                        1162
```

Fig. 3. Alignments of the peplomer sequences of FIPV (strain 79-1146), IBV (strain M41) and MHV (strain A59). Tentative or arbitrary alignments are indicated by extra spacing between sequence lines. Identical residues are underlined. Thke cleavage sites in the peplomer precursor proteins of MHV and IBV are indicated by arrowheads (Cavanagh et al., 1986; L. Sturman, pers, commun.), the presumptive signal peptides and transmembrane anchors by black bars.

INDICATIONS FOR COILED α-HELICES IN THE STALK OF THE PEPLOMER

The three-dimensional structure of the coronaviral peplomer has not yet been determined. Therefore, no data are available on how monomers interact to form a stable multimer. Also, the molecular basis for the typical elongated appearance of the peplomers is unknown. In other elongated protein molecules, like the haemagglutinin of influenza virus and reovirus, long α-helices of the monomers interlock in a coiled coil (Wilson et al., 1981; Bassel-Duby et al., 1985). As reviewed by Cohen and Parry (1986) this structure stabilizes the multimer and imparts the elongated character to the molecule. Indicative for α-helices forming a coiled coil is a seven-residue 'heptad' repeat in the amino acid sequence (a, b, c, d, e, f, g), in which the residues in the positions a and d generally have an apolar character. In the α-helix, these residues are aligned, resulting in a continuous hydrophobic stretch along the axis. Such hydrophobic 'backbones' form the interface between interlocking helices.

Conceivably, a similar structure could also be present in the peplomers of coronaviruses. Systematic examination of the peplomer sequences revealed the presence of two heptad repeats in the C-terminal region of E2 (Fig. 4). The repetitious character of these regions is also obvious in the Diagon plot (Fig. 1). One of the repeats is located immediately upstream of the transmembrane anchor. The presence of this repeat (residues 1328-1380, 1055-1080 and 1214-1251 of the peplomer proteins of FIPV, IBV and MHV, respectively) is well conserved, in spite of the low degree of amino acid conservation in this region.

The other heptad repeat is even longer and located further upstream (Fig. 4; residues 1067-1149, 796-864 and 972-1041 of the E2 proteins of FIPV, IBV and MHV ,respectively). In Fig. 5, this repeat is visualized by listing the sequence in alternating rows of four and three residues (a 'helical net'). As indicated in Fig. 4, both heptad repeats coincide in all three proteins with regions devoid of helix-breaking proline residues.

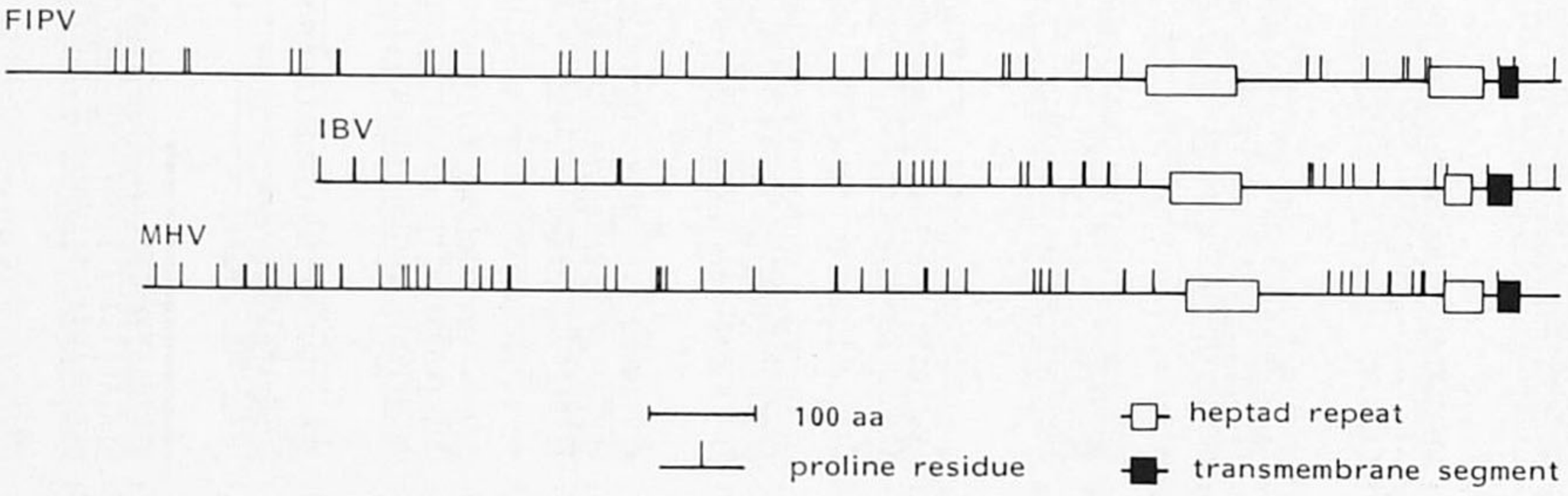

Fig. 4. Indications for long α-helices in the peplomer protein.

The two heptad repeats are indicative for the presence of two α-helices in the peplomer protein that are capable to interlock with other helices. For the major repeat a helix with a length of at least 10 nm can be predicted. This helix would be longer than the longest helix of influenza virus haemagglutinin (7.5 nm) and would extend over about half the length of the peplomer (17-20 nm). It is tempting to speculate that, as in the case of the haemagglutinins, such an α-helix interlocks with the α-helix(-ces) of (an)other mononomer(s) to form a coiled coil. This structure could stabilize the multimer and account for the characteristic elongated appearance of the stalk of the coronavirus peplomer.

ACKNOWLEDGMENTS

This research was supported by a grant from Duphar B.V., Weesp, The Netherlands.

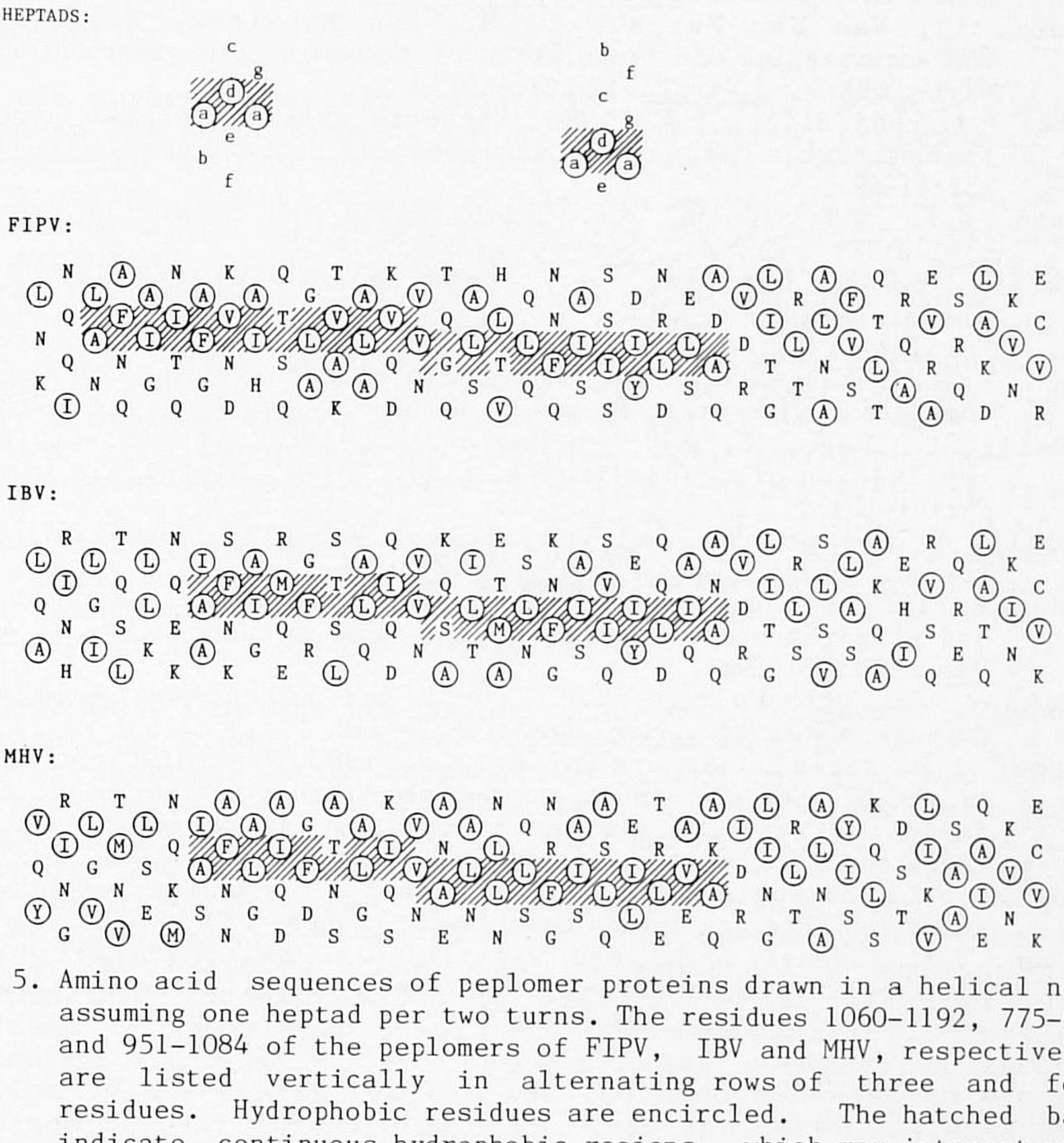

Fig. 5. Amino acid sequences of peplomer proteins drawn in a helical net, assuming one heptad per two turns. The residues 1060-1192, 775-907 and 951-1084 of the peplomers of FIPV, IBV and MHV, respectively, are listed vertically in alternating rows of three and four residues. Hydrophobic residues are encircled. The hatched bars indicate continuous hydrophobic regions, which may interact with other α-helices.

REFERENCES

Bassel-Duby, R., Jayasuriya, A., Chatterjee, D., Sonenberg, N., Maizel Jr, J.V., and Fields, B.N., 1985, Sequence of reovirus haemagglutinin predicts a coiled-coil structure, Nature, 315:421-423.

Binns, M.M., Boursnell, M.E.G., Cavanagh, D., Pappin, D.J.C., and Brown, T.D.K., 1985, Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV, J. gen. Virol., 66:719-726.

Boyle, J.F., Pedersen, N.C., Evermann, J.F., McKeirman, A.J., Otts, R.L., and Black, J.W., 1984, Plaque assay, polypeptide composition and immunochemistry of feline infectious peritonitis virus and feline enteric coronavirus isolates, Adv. Exp. Med. Biol., 173:133-147.

Cavanagh, D., 1983, Coronavirus IBV: structural characterization of the spike protein, J. gen. Virol., 64:2577-2583.

Cavanagh, D., Davis, P.J., Pappin, D.J., Binns, M.M., Boursnell, M.E.G. and Brown, T.D.K., 1986, Coronavirus IBV: partial amino terminal sequencing of spike polypeptide S2 identifies the sequence Arg-Arg-Phe-Arg-Arg at the cleavage site of the spike precursor propolypeptide of IBV strains Beaudette and M41, Virus Res., 4:133-143.

Cohen, C., and Parry, D.A.D., 1986, -Helical coiled coils - a widespread motif in proteins, Trends Biochem. Sci., 11:245-248.

Jacobs, L., Van der Zeijst, B.A.M., and Horzinek, M.C., 1986, Characterization and translation of transmissible gastroenteritis virus mRNAs, J.Virol., 57:1010-1015.

Kapke, P.A., and Brian, D.A., 1986, Sequence analysis of the porcine transmissible coronavirus nucleocapsid protein gene, Virology, 151:41-49.

Lipman, D.J., and Pearson, W.R., 1985, Rapid and sensitive protein similarity searches, Science, 227:1435-1441.

Niesters, H.G.M., Lenstra, J.A., Spaan, W.J.M., Zijderveld, A.J., Bleumink-Pluym, N.M.C., Hong, F., van Scharrenburg, G.J.M., Horzinek, M.C. and van der Zeijst, B.A.M., 1986, The peplomer protein sequence of the M41 strain of coronavirus IBV and its comparison with other Beaudette strains, Virus Res., 5:253-263.

Siddell, St., Wege, H., and Ter Meulen, V., 1982, The structure and replication of coronaviruses, Curr. Topics Microbiol. Immun., 99:131-163.

Siddell, St., Wege, W., and Ter Meulen, V., 1983, The biology of coronaviruses, J. gen. Virol., 64:761:776.

Staden, R., 1982, Diagon: an interactive graphics program for comparing and aligning nucleic acid or amino acid sequences, Nucl. Acid. Res., 10:2951-2961.

Sturman, L.S., and Holmes, K.V., 1983, The molecular biology of coronaviruses, Adv. Virus Res., 28:35-112.

Sturman, L.S., Ricard, C.S., and Holmes, K.V., 1985, Proteolytic cleavage of the E2 glycoprotein of murine coronavirus: activation of cell-fusion activity of virions by trypsin and separation of two different 90K cleavage fragments, J. Virol., 56:904-911.

Wickner, W.T. and Lodish, H.F., 1985, Multiple mechanisms of protein insertion into and across membranes, Science, 230:400-407.

Wilson, I.A., Skehel, J.J., and Wiley, D.C., 1981, Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution, Nature, 289:366-373.